PB-229 354

GUIDELINES FOR THE DEVELOPMENT OF AN AIR QUALITY DATA SYSTEM

Charles Zimmer, et al

Pedco-Environmental

Prepared for:

Environmental Protection Agency

September 1973

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

PB 229 354

BIBLIOGRAPHIC DATA SHEET	1. Report No. EPA-450/3-73-008	2.

4. Title and Subtitle "Guidelines for the Development of An Air Quality Data System"	5. Report Date September 1973 6.
7. Author(s) Pedco Environmental	8. Performing Organization Rept. No.
9. Performing Organization Name and Address Pedco-Environmental Specialists Suite 8 Atkinson Square Cincinnati, Ohio 45246	10. Project/Task/Work Unit No. 11. Contract/Grant No. Contract No. 68-02-0044
12. Sponsoring Organization Name and Address EPA,OAQPS,MDAD,NADB Research Triangle Park, N. C. 27711	13. Type of Report & Period Covered Final 1/11/72-Present 14.

15. Supplementary Notes

16. Abstracts

This report defines the steps to take in analyzing aerometric data requirements and defining a data handling system. It illustrates various decisions which were made and the reasons for them in the data handling system of the city of Cincinnati. It includes the steps which are necessary to computerize the system and designing input and output formats. Files are addressed briefly with a general description of file types and media.

17. Key Words and Document Analysis. 17a. Descriptors

Management Information System
Computers
ADP (Automatic Data Processing)
Air Quality Data System
Air Pollution
Ambient Air Data
Aerometric Data
Sysyem
Guideline

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group 13B

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Release Unlimited	UNCLASSIFIED	77
	20. Security Class (This Page) UNCLASSIFIED	22. Price 7.00 / 1.45

EPA-450/3-73-008

GUIDELINES FOR THE DEVELOPMENT OF AN AIR QUALITY DATA SYSTEM

by

Charles Zimmer, Eugene Forte and Robert Braley

Pedco-Environmental
Suite 8 Atkinson Square
Cincinnati, Ohio 45246

Contract No. 68-02-0044

EPA Project Officer: Gerald Nehls

Prepared for

ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Water Programs
Office of Air Quality Planning and Standards
Research Triangle Park, N. C. 27711

September 1973

This report is issued by the Environmental Protection Agency to report technical data of interest to a limited number of readers. Copies are available free of charge to Federal employees, current contractors and grantees, and nonprofit organizations - as supplies permit - from the Air Pollution Technical Information Center, Environmental Protection Agency, Research Triangle Park, North Carolina 27711, or from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22151.

This report was furnished to the Environmental Protection Agency by Pedco-Environmental, Suite 8 Atkinson Square, Cincinnati, Ohio, in fulfillment of Contract No. 68-02-0044. The contents of this report are reproduced herein as received from Pedco-Environmental. The opinions, findings, and conclusions expressed are those of the author and not necessarily those of the Environmental Protection Agency. Mention of company or product names is not to be considered as an endorsement by the Environmental Protection Agency.

Publication No. EPA-450/3-73-008

ACKNOWLEDGEMENT

Many individuals and organizations have been helpful in developing this guideline; for their contributions the project management extends its sincere gratitude.

The contributions of Messrs. Eugene Ermenc and Robert Shaw of the City of Cincinnati Air Pollution Control Division, Robert Callahan and Justin Fisher of the City of Cincinnati and Hamilton County Regional Computer Center were of particular significance.

Mr. Gerald Nehls, Environmental Protection Agency and Mr. Charles Schumann, City of Cincinnati Air Pollution Control Division served as project officers for their respective organizations. Mr. Charles Zimmer, PEDCo-Environmental Specialists, Inc., the project manager, was assisted by Eugene Forte and Robert Braley.

This report was furnished to the Environmental Protection Agency by PEDCo-Environmental Specialists, Inc. in fulfillment of a contract with the City of Cincinnati, Ohio which was supported in part by funds contributed to the City of Cincinnati, Ohio by the Environmental Protection Agency. The contents of this report are reproduced herein as received from the contractor. The opinions, findings, and conclusions are those of the authors and are not necessarily those of the Environmental Protection Agency. Mention of company or product names does not constitute endorsement by the Environmental Protection Agency.

TABLE OF CONTENTS

		PAGE
1.0	INTROLUCTION	1
1.1	Background	1
	1.1.1 Requirements of the Clean Air Act	1
	1.1.2 Cincinnati Experience	2
1.2	Purpose of this Report	3
	1.2.1 System Guideline Document	3
	1.2.2 Definition of Future Data Acquisition Requirements	3
	1.2.3 Personnel Requirements	3
2.0	SURVEY DATA HANDLING REQUIREMENTS	5
2.1	Existing Procedures	5
	2.1.1 Manual Procedures	5
	2.1.2 Automated Procedures	5
	2.1.3 Data Handling - Cincinnati	6
2.2	Constraints Imposed by State and Federal Data Systems	6
2.3	Interstate Agreements	8
3.0	DESIGN INPUT DATA FORMATS	9
3.1	Coding	9
	3.1.1 Station Identification	9
	3.1.2 Parameter Identification	11
	3.1.3 Monitoring Period	13
3.2	Data Recording Formats	13
	3.2.1 Continuous Monitoring	16
	3.2.2 Intermittent Data	19
3.3	Cincinnati's Data Input Formats	24
	3.3.1 Continuous Monitors	24
	3.3.2 Intermittent Data	31

TABLE OF CONTENTS
(continued)

		PAGE
4.0	OUTPUT DATA FORMATS	34
4.1	Survey Users Requirements	34
4.2	Clarity of Content	35
4.3	Cincinnati Formats	35
	4.3.1 Data Listings	35
	4.3.2 Monthly Report	37
	4.3.3 Summary Report of Intermittent Data	44
	4.3.4 Data Analysis	45
	4.3.5 Submit Data to NADIS	45
5.0	DATA STORAGE	53
5.1	Storage Media	53
	5.1.1 Data Record Forms	53
	5.1.2 Punched Paper Tape	53
	5.1.3 Punched Cards	53
	5.1.4 Magnetic Tape	54
	5.1.5 Direct Access Storage	55
	5.1.6 Selection of a Storage Media	55
5.2	Information Management	56
	5.2.1 Data Storage	57
	5.2.2 Record Formats	58
5.3	Structure of the Cincinnati Files	59
	5.3.1 Use of Secondary Storage Media	59
	5.3.2 Record Format - Disk File	59
	5.3.3 Record Format	68

1.0 INTRODUCTION

1.1 BACKGROUND

1.1.1 Requirements of the Clean Air Act

The Clean Air Act of 1970 has had a profound effect upon all aspects of local, State, and Federal air pollution control programs. Following the adoption of National Ambient Air Quality Standards for selected air contaminants, each State has submitted its plan for the implementation of these Standards. The availability of aerometric data has been and will continue to be a very critical part of each implementation plan.

Information relative to existing air quality and meteorological conditions is of paramount importance in the determination of specific emission control regulations required to insure the attainment of the air quality standards. Air quality data are essential for the evaluation of the effectiveness of such regulations. Once the air quality standards have been achieved, air quality data will provide the information base needed for maintenance of these standards.

Most State and local control agencies anticipate a significant expansion in their air quality and meteorological monitoring programs. Expansion of these programs will greatly increase the quantity of data to be handled. As a result, most

agencies will find it necessary to expand existing, and develop new, data handling systems. All of the State agencies and many local agencies will find it necessary to use computer oriented data storage and retrieval systems.

1.1.2 Cincinnati Experience

The Ohio Air Pollution Control Board has delegated to the City of Cincinnati, Division of Air Pollution Control (DAPC), responsibility for providing most control agency functions to the four-county Ohio portion of the Cincinnati Interstate Air Quality Region. Included among these responsibilities is that of air quality and meteorological monitoring. The DAPC is responsible for the installation, operation, and maintenance of all monitoring stations. By 1975 this network will be comprised of 40 stations, of which as many as nine will be equipped with one or more continuous pollutant and/or meteorological monitoring devices. The balance of the stations will be equipped with one or more devices for collecting 24-hour pollutant samples.

In line with this expansion of aerometric monitoring activities, the DAPC has recently designed, and is now implementing, a computer oriented data handling system. This automated system will provide the data summaries and statistical analyses required by management to carry out an effective control program.

1.2 PURPOSE OF THIS REPORT

1.2.1 System Guideline Document

In the development of the Cincinnati Aerometric Data Handling System, many problems had to be solved and decisions had to be made by management, relative to the scope and complexity of the system. This experience provides a body of information which can be of significant value to other control agencies. This document is intended to serve as a guideline for the design of aerometric data storage and retrieval systems.

1.2.2 Definition of Future Data Acquisition Requirements

At the present time, most control agencies are not completely aware of the importance of aerometric data in their day-to-day activities. Because of this, many agencies have not formulated a policy on the use of real-time data acquisition systems. It is most important that a control agency avoid "overkill" in its approach to data handling. To this extent, the approach taken by Cincinnati is to develop a data handling system which operates in a batch mode, that is to say data are physically taken from the field monitoring station to the laboratory and then to the computer. Once this system is operational, it may be expedient to incorporate a real-time link between the continuous monitoring device and the computer. This additional level of complexity involving telemetry can easily be incorporated into the data system.

1.2.3 Personnel Requirements

The design and implementation of an aerometric data handling system requires specialized personnel resources. It is

unlikely that most control agencies will have such individuals on their staffs. On the other hand, most control agencies have access to a computer facility operated by some other entity within the government. The systems analysts and computer programmers available from a central computer services group will find it extremely difficult to design the data system without the assistance of air pollution specialists. Thus, most agencies will find it necessary to designate an air pollution specialist to work with the computer systems analyst and programmer. Thus, in effect, a team approach is used in the design and implementation of the system.

2.0 SURVEY DATA HANDLING REQUIREMENTS

2.1 EXISTING PROCEDURES

The first step in the design of an aerometric data system is to review the data handling procedures now in use by the agency. This review should provide the details of data recording formats, coding techniques, method of data storage, and the types of reports which are prepared from the data. Additionally, this review should include an estimate of the volume of data being handled by the system.

2.1.1 Manual Procedures

Historically, many control agencies have been collecting limited amounts of 24-hour pollutant concentration data. Some agencies may also obtain 2-hour soiling index measurements from one or two stations. For the most part, the control agencies have tended toward manual procedures for this type of data. Preparation of a data flow diagram, and documentation of input and output formats used to handle such data, is generally quite easy.

2.1.2 Automated Procedures

It is most important that existing automated procedures be carefully reviewed. Where automated procedures are being used, input data formats, coding techniques, record formats, file

structures and data retrieval formats may already be well established. These existing procedures should not necessarily constrain the design of the total aerometric data system. However, where practical, they should be incorporated into the new system.

2.1.3 Data Handling - Cincinnati

Air quality data has been collected in Cincinnati for 25 years. This data has resulted from programs supported entirely by the resources of the local agency and from cooperative studies with EPA (and its predecessors). The EPA studies included such long term air monitoring activities as the National Air Sampling Network, the National Gas Sampling Network, the Continuous Air Monitoring Program, and related studies. For the most part, the local agencies have not been required to develop data handling procedures for the EPA cooperative studies.

The volume of data resulting from locally supported programs was of such quantity that manual data handling was adequate. The local agency developed data recording forms for measurements of monthly dust fall, weekly wind blown particulates, 24-hour suspended particulates and gases, and 2-hour soiling index. No special numerical coding was developed which in any way imposed a constraint on the development of the automated data handling system.

2.2 CONSTRAINTS IMPOSED BY STATE AND FEDERAL DATA SYSTEMS

Local control agencies are required to submit air quality data to the State agency. Likewise, Federal regulations require the State agency to submit such data to EPA. The requirements for such data should be carefully considered in the design of a

data handling system. It is generally true that the agency responsible for the operation of the air sampling network requires the most detail, in so far as data is concerned. Thus, with a local agency, if its own requirements are satisfied, there should be no trouble in supplying the level of detail required by the State and EPA.

The record format and coding used by EPA need not necessarily be used in the design of a local or State data system. The important thing is that the data input include all of the data elements required by the higher echelon. Fundamentally, the data should be identified as follows:

Agency Collecting the Data

Project Code, i.e., routine sampling or special sampling

Location of Sampling Station

Pollutant Name

Method of Sampling and Analysis

Units of Recording Measurements

Sample Averaging Time, (i.e. 1 hr, 2 hr, etc.)

Data and Time of Sample

Decimal Point Location

Actual Data Values

Included in the aerometric data system would be the necessary information to transform data elements to the coding specified by EPA (or the State).

2.3 INTERSTATE AGREEMENTS

Within Interstate Air Quality Control Regions, it is necessary to interchange data between control agencies in two or more States. Representatives of all States involved in an Interstate AQCR should meet and discuss their mutual requirements for data. If at all feasible, an attempt should be made to standardize on station identification and coding techniques.

3.0 DESIGN INPUT DATA FORMATS

3.1 CODING

As was previously noted, each data element entered in a data storage and retrieval system must be uniquely identified. With a manual system, all pertinent information can be entered on a data form which is then placed in a filing cabinet. When a computer oriented data system is used, it is impractical to store all identifying information in its original form. In order to maintain efficiency and minimize the cost of data input and storage, the number of characters of identifying information should be kept to a minimum. This can be accomplished with the use of numerical codes for descriptive information such as station identification, method of determination, units of measurement, date of collection, etc.

It is worthy of note at this point, that while coding is efficient in terms of the computer, coding is most inefficient from the point of view of the user. Thus, it is always desirable to restore information to its original form when data is retrieved from the computer file.

3.1.1 Station Identification

A numerical code is most often used for station identification. The structure of this code is dependent upon the way

in which data must be locateable for retrieval by the user. In its most simple form the individual stations are numbered from 1 to N, with the value of N sufficiently large to accommodate some reasonably expected increase in the air monitoring activities. For most local agencies a two-digit code is sufficient since up to 99 stations are possible. Station numbers are assigned sequentially as new ones are added.

As the scope of an aerometric monitoring network expands, it is usually desirable to retrieve data by various subsets of stations. For example, a state agency may maintain a data bank consisting of data from a large number of political jurisdictions. In this situation, it may be necessary to use a hierarchial code in which Air Quality Control Region, City and/or County, as well as sampling site, is identified, such as:

AQCR	xx
City/County	xx
Station No.	xx

A hierarchial code is particularly useful in retrieving data from the file. To retrieve data for a given sampling site, the AQCR, City/County, and Station numbers are specified. To retrieve the data for all sampling sites in a city, specify the AQCR and city numbers. To retrieve data for all cities/counties in an AQCR, specify only the AQCR number. Finally, to retrieve all data in the file, the AQCR, City/County, and station numbers are all left blank (or insert zero's).

A nine (9) digit is used by SAROAD for station identification as follows:

State	xx
Area (City or County)	xxxx
Site	xxx

The SAROAD station coding used by EPA is necessarily different from that which may be desirable for use by a State or local air pollution control agency. Because State and local agencies are required to submit air quality data to EPA, it is desirable to include in the data retrieval programs, the ability to convert station identification codes to the SAROAD code.

It is necessary to exercise judgement in the development of the station I.D. code which is used with data being entered into the system. As an example, it may appear to be desirable to develop the station I.D. code on the basis of grid coordinates. To do so with the Universal Transverse Mercator System (UTM) may require as many as 17 digits for zone as well as Easting and Northing coordinates. Since the station I.D. must be included as part of the input with new data, the number of characters used should be minimized.

3.1.2 Parameter Identification

The parameter identification must include; the parameter name, method of determination, units of measurement, and decimal point location.

3.1.2.1 Parameter Name - The code for parameter can be either a sequential number or a hierarchial number code. Since most control agencies routinely collect data on no more than 10-15

parameters, a sequential two-digit code may be desirable. If it becomes necessary to structure the computer files according to either meteorological or air quality parameters, and the air quality parameters are further classified in some manner, it may be necessary to use a hierarchial number code. SAROAD employs a five-digit code to provide maximum flexibility in the assignment of parameter codes. This code provides the ability to retrieve data for individual pollutants or selected categories of pollutants with a single request to the data base.

Again it is desirable to include in the computerized system a procedure for converting the parameter code to the SAROAD code.

3.1.2.2 Method of Determination - For many air pollutants there are several methods of sample collection and analysis currently being used. In some instances, there is sufficient evidence to indicate that the results from the several methods are not directly comparable. Because of this, it is desirable to develop a code for method of determination. For the most part, a one-digit numerical code is sufficient to handle the number of different methods being used by an agency for a given pollutant. The SAROAD code is two digits to maintain adequate flexibility.

3.1.2.3 Units of Measurement - When data is retrieved from a computerized storage media, it is important that units of measurement be included on the printed copy. The units of measurement can be stored with the data element or be included as a table of constants in each computer program which is written to retrieve data. If the units of measurement are stored with the

data, it is necessary that it be included with the input data. Because of the extreme flexibility in data input desired for SAROAD, a two-digit code is used by EPA (Table 1).

3.1.3 Monitoring Period

3.1.3.1 Date and Time - The monitoring period is identified by date and time of day. It is recommended that the beginning of the period be recorded by year, month, day, hour, and minute. For monitoring periods of one hour or more, it is good to begin on the hour, in which case it is not necessary to record minutes. Since many agencies use the period midnight-to-midnight for 24-hour measurements, the beginning hour can also be deleted.

When a data acquisition system is used with continuous monitoring equipment, it is expedient to record year, Julian day, hour and minutes at the beginning of the monitoring period. With most computer installations, algorithms to convert from Julian day to month and day are readily available.

3.1.3.2 Time Interval - The time interval over which a specific parameter is measured must be associated with each data element. This can be accomplished in two ways; (1) include a code for time interval with the identification information for each data element, and (2) structure the file by time interval so that all 24-hour data is maintained in one file, 2-hour data in another file, etc. The time interval code used for the submission of data in SAROAD format is shown in Table 2.

3.2 DATA RECORDING FORMATS

Aerometric data is generated in two basic ways; (1) with continuous monitoring devices, and (2) with intermittent monitoring devices. The output from a continuous monitor is

TABLE 1. UNITS

Code number	Units
01	micrograms/cubic meter (25° C, 1013 millibars)
02	micrograms/cubic meter (0° C, 1013 millibars)
03	nanograms/cubic meter (25° C, 1013 millibars)
04	nanograms/cubic meter (0° C, 1013 millibars)
05	milligrams/cubic meter (25° C, 1013 millibars)
06	milligrams/cubic meter (0° C, 1013 millibars)
07	parts per million (volume/volume)
08	parts per billion (volume/volume)
09	COHS/1000 linear feet
10	RUDS/10,000 linear feet
11	meters/second
12	miles/hour
13	knots
14	degrees
20	microns
30	picocuries/cubic meter
31	microcuries/cubic meter
32	picocuries/square meter
33	microcuries/square meter
34	picocuries/cubic centimeter
35	picocuries/gram
50	number of threshold levels
70	milligrams F/100 square centimeters-day
80	milligrams SO_3/100 square centimeters-day
81	micrograms SO_2/square meter-day
90	tons/square mile-month[a]
91	milligrams/square centimeter-month[a]
92	micrograms/cubic meter-month[a]
98	milligrams SO_4^{2-}/square centimeters-30 days
99	milligrams/square centimeters-30 days

[a]On a calendar-month basis.

Source: SAROAD Users Manual
Office of Air Programs Publication
NO. APTD-0663

TABLE 2. TIME INTERVAL

Code	Data observed over a period of:
1	1 hour
2	2 hours
3	4 hours
4	6 hours
5	8 hours
6	12 hours
7	24 hours
8	1 month
9	3 months
A	1 week
B	3 hours
C	Composite data
D-Z	For future expansion

Source: SAROAD Users Manual
Office of Air Programs
Publication No. APTD-0663

recorded either in the field on a strip chart, paper tape, or magnetic tape, or is telemetered to a central station for immediate computer processing. With intermittent monitoring devices, a sample is collected in the field and then sent to a laboratory for analysis. Typically, results from the laboratory are recorded on an appropriate form for subsequential data handling.

3.2.1 Continuous Monitoring

3.2.1.1 Strip Chart Data Reduction - The reduction of data from strip charts can be done entirely manually or semi-automatically with the use of a chart reader. When charts are read manually, hourly measurements are "eye-balled" and the results recorded in engineering units on a data record form. A very convenient form for recording hourly data is the one used by SAROAD (Figure 1). This form provides for all the necessary identification and permits recording up to 4 digits for each parameter. The form is designed for direct keypunching and is sufficiently flexible to handle a wide range of parameters from many different monitoring stations.

A chart reader can be used to minimize the manual effort involved in reading strip charts. Typically, hourly data values are reduced from the charts and automatically printed by an on-line typewriter; or preferably, entered through a keypunch on to punched cards. In either case, the output from the chart reader can be formatted similarly to that shown in Figure 1.

3.2.1.2 Data Acquisition Systems - The inclusion of a data acquisition system in a continuous monitoring program can minimize the manpower requirements for data handling. Output from

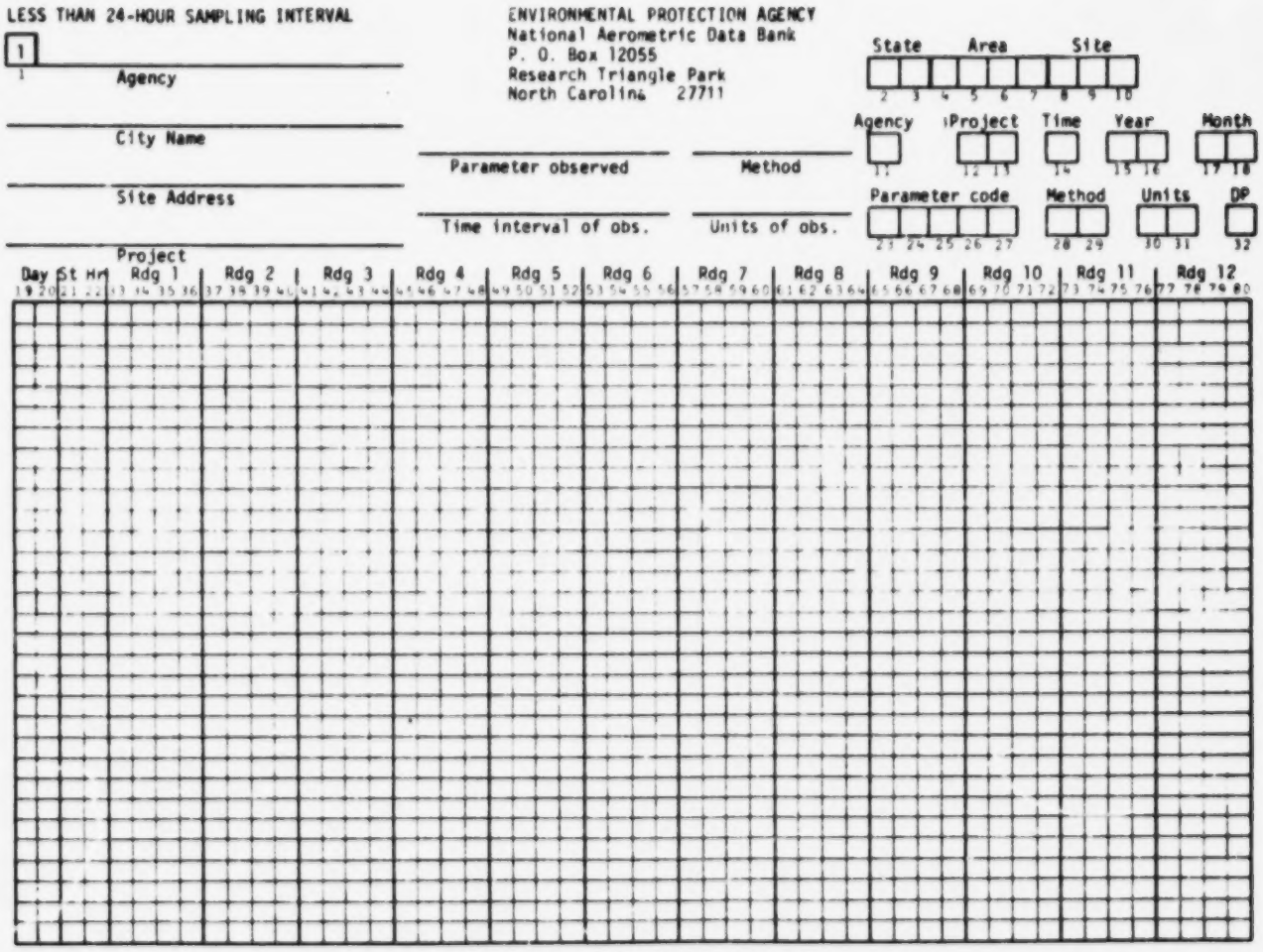


Figure 1. SAROAD Hourly Data Form.

the data acquisition system may be recorded on an intermediate storage media, or processed on-line by a computer.

A data acquisition system digitizes the analog signal from the sensor, usually converts the data element to engineering units and records the data on tape, or transmits it to a computer. When there are a number of sensors at a station, the data from all sensors may be multiplexed and assembled into single message units. Included in such a message unit is the identification information for the station and the sensor, as well as code indicating the status of each sensor.

There are two ways in which the data acquisition system can handle the analog signal from a sensor; (1) record instantaneous pulses, and (2) through the use of an electronic integrator, record time-averaged values. The frequency of recording instantaneous values and the length of time for integration is determined by the way in which the data are used in decision making and planning. Most agencies find that data values averaged over 30 to 60 minutes satisfy most of their requirements.

Current practice by most agencies is to record instantaneous data values and use the computer to determine time-averaged values. Basically, this makes the computer the integrator. The frequency of recording instantaneous data values determines the precision with which a time-averaged value can be computed. Frequencies of recording now being used range from one per minute to four per hour. A recording frequency of 12 per hour (every 5 minutes) is used for the Continuous Air Monitoring Program (CAMP). When hourly averages are of primary concern, the recording of instantaneous values at 5-minute intervals is a

ed compromise between maintaining reasonable precision of the average value and the cost of data handling.

3.2.1.3 Station Activity Record - With continuous monitoring devices, it is important that all pertinent information regarding station operation be available to the individual responsible for data validation. When data is reduced from strip charts, it is convenient for the station operator to enter comments directly onto the chart. With an automated data system, the operator must have another mechanism for transmitting information to the computer. A most convenient way to accomplish this is through the use of an operator's log. An example of such a log is the one used by CAMP (Figure 2). The log provides a means to invalidate data for periods when it is known that a sensor was malfunctioning. Also, the operator can record necessary information concerning instrument zero and span checks, etc.

3.2.2 Intermittent Data

Intermittent data is primarily associated with air quality parameters. Typically, a sampling media (filter paper or reagent) is prepared by the laboratory and sent to the field station where it is exposed for the appropriate time. After the sampling period, the media is returned to the laboratory for analysis. Finally, the pollutant concentration is determined and recorded for subsequent use.

A record must be maintained throughout the period from media preparation to recording the pollutant concentration. Initially, the laboratory determines and records the weight of each filter for the hi-volume sampler. In the field, the operator records the necessary station identification, start

Figure 2

SEC 468
(REV 10-61)

RECORD OF OPERATOR'S LOG

DATE: _____ to _____
OPERATOR _____
GAS _____ (6 7)

STATION _____ (1 2 3 4 5)

MO.	DA.	YR.	D*	ITEM	START	STOP	PURGE	COMMENTS
8 9	10 11	12 13	14	15 16 17	18 19 20 21	22 23 24 25	27 28 29 30	

* Day of Week.

time, stop time, and air flow rate. A typical example of a field sample form is shown in Figure 3. Finally, the laboratory performs the requisite analysis and computes the pollutant concentration.

The output from the laboratory can be in one of two forms; (1) pollutant concentration in engineering units, or (2) intermediate results prepared for computer calculation of concentration.

3.2.2.1 Pollution Concentration in Engineering Units - The laboratory analyst performs all necessary calculations to determine the concentration of the parameter in micrograms per cubic meter or other appropriate units. Results are verified to insure their reliability, and the final data value is recorded on an appropriate form.

If the data are to be entered into a computer storage and retrieval system, the data record form should be designed for computer input. An example of a form designed for direct keypunching is the one used with SAROAD (Figure 4). If an agency has the necessary hardware available at its computer center, the need for keypunching can be eliminated through the use of optical character recognition forms and readers. It is recommended that the computer center be consulted prior to selecting the method to be used for computer data input.

3.2.2.2 Laboratory Reports Intermediate Results - As in other situations involving highly repetitive tasks, some functions of the laboratory can be automated if the volume of work is sufficiently large. For the purpose of this report, only automation

SAMPLE NO._____

STATION NO._____

	START OF SAMPLING	FINISH OF SAMPLING
DATE	_____	_____
TIME	_____	_____
FLOW RATE	_____	_____
DATE	_____	_____
FILTER WEIGHT	_____	_____
ROOM TEMPERATURE	_____	_____
RELATIVE HUMIDITY	_____	_____

TOTAL VOLUME SAMPLED _____ CUBIC METERS

TOTAL WEIGHT GAINED _____ MILLIGRAMS

DUST LOADING _____ μGMS/M^3

STATION OPERATOR_____

WEIGHING OPERATOR_____

BAAPCD
WS:fm
10/8/70

Source: Bay Area Air Pollution Control District

Figure 3. Particulate sampling record.

Figure 4

ENVIRONMENTAL PROTECTION AGENCY
National Aerometric Data Bank
Research Triangle Park, N. C. 27711

SAROAD Daily Data Form

OMB No. 158-R0012
Approval expires 6/30/76

24-hour or greater sampling interval

[2] (1)

Agency

City Name

Site Address

Project Time Interval

State (2 3) Area (4 5 6 7) Site (8 9 10)

Agency (11) Project (12 13) Time (14) Year (15 16) Month (17 18)

Day (19 20)	St (21)	Hr (22)	Name PARAMETER Code (23 24 25 26 27) Method (28 29) Units (30 31) DP (32) (33 34 35 36)	Name PARAMETER Code (37 38 39 40 41) Method (42 43) Units (44 45) DP (46) (47 48 49 50)	Name PARAMETER Code (51 52 53 54 55) Method (56 57) Units (58 59) DP (60) (61 62 63 64)	Name PARAMETER Code (65 66 67 68 69) Method (70 71) Units (72 73) DP (74) (75 76 77 78)
0	1					
0	2					
0	3					
0	4					
0	5					
0	6					
0	7					
0	8					
0	9					
1	0					
1	1					
1	2					
1	3					
1	4					
1	5					
1	6					
1	7					
1	8					
1	9					
2	0					
2	1					
2	2					
2	3					
2	4					
2	5					
2	6					
2	7					
2	8					
2	9					
3	0					
3	1					
DP →			4 3 2 1 0	4 3 2 1 0	4 3 2 1 0	4 3 2 1 0

of the routine computations to determine pollutant concentrations is treated.

When the laboratory analyst completes the analytical work, it is necessary to perform some rather simple arithmetic to determine pollutant concentration. These calculations can be handled with extreme speed on a computer. With the appropriately designed data form, the field operator initiates the transaction by recording the necessary station identification, date, sampling time, air flow rates, etc. In the laboratory, the analyst records the proper weight, meter reading, etc. Next, the form is keypunched, and finally the data is processed by the computer. The computer enters data into a file and prepares a listing of all transactions which can be used in data validation. Finally, the laboratory validates the computed data and prepares the necessary transaction to change or delete invalid data which has already been stored in the computer file.

Before attempting to initiate this procedure, careful thought should be given to the following:

1. Additional steps are being added which take time to execute, introduce the need for additional checking, and increase the elapsed time in processing samples.

2. Special computer programs must be written and maintained.

3. There is a continuing cost for the computer.

4. Relieving the analyst of the responsibility for the final result may introduce a lack of concern over the validity of results.

3.3 CINCINNATI'S DATA INPUT FORMATS

3.3.1 Continuous Monitors

Data acquisition hardware is included at each continuous

monitoring station. The system is designed for a maximum of 16 air quality or meteorological sensors. Instantaneous values for each parameter are recorded on magnetic tape at 5-minute intervals. Data values plus pertinent identification information are written in a 156 character record as shown in Figure 5.

The input format permits a maximum of 99 stations. Each station may have from 1 to 16 parameter sensors. In the present design, the system requires that the parameters be recorded in the same order at each station. Because of the fixed record length of 156 characters, the system automatically writes no data codes for sensors which are not operative at a given station.

The Station Status Code (Ref. No. 4 - Figure 5) provides a means of checking the validity of the signal from each sensor. Under normal operating conditions this status code = Ø. Once each day the system automatically checks the low calibration (10% of scale), Code = 1, and the high calibration (70% of scale), Code = 2, for each sensor. This information is used by the computer as a validity check of the signal from each sensor. Should either the low or high calibration differ by more than 5 percent from its correct value, the computer is programmed to invalidate the data for that parameter for the preceding 24 hours.

Under normal operation (Station Status Code = Ø) a two-point validity check is made on the electronics of the data acquisition system each time a record is written on tape. The use of a two-point check (i.e. ZERO and SPAN) provides reasonable assurance that a drift in the electronics, which would otherwise result in spurious data, will not go undetected. The DVM Zero Check (Ref. No. 5) is set to emit a constant + ØØØØ and the DVM Span Check

Figure 5. File Layout - Field Magnetic Tape

PAGE 1

X	INPUT		OUTPUT	SYSTEM NAME: Continuous Data	DATE
	REFERENCE		INTERMEDIATE		
PROGRAM NUMBER(S)				FILE NAME: Raw Data	RECORD NAME: Five Minute Data
SEQUENCE				FREQUENCY	DISPOSITION	NO OF RECORDS: PEAK: NORMAL:
DEVICE DESCRIPTION: Magnetic Tape				PRINT/PUNCH DOCUMENT SIZE: WIDE X LONG	NO OF COPIES
REMARKS					

REF NO	DATA ELEMENT	NUMBER OF CHARS	DEC	BYTES	A AN P B	SOURCE	RELATIVE LOCATION
1	Station No.	2			N	Ø1-Ø9	1-2
2	Date	5			N		3-7
	Year	2			N	>72	3-4
	Day	3			N	ØØ1-366	5-7
3	Time	4			N		8-11
	Hour	2			N	ØØ-23	8-9
	Minutes	2			N	ØØ-59	10-11
4	Station Calibrate Status	1			N	Ø,1,2	12
5	DVM Zero Check	8			N		13-20
	Channel Status	1			N	Ø,1,2,3	13
	Channel ØØ	2			N	Always ØØ	14-15
	Polarity	1			N	+ or -	16
	Value	4			N		17-20
6	DVM Span Check	8			N		21-28
	Channel Status	1			N	Ø,1,2,3	21
	Channel Ø1	2			N	Always Ø1	22-23
	Polarity	1			N	+ or -	24
	Value	4			N		25-28
7	Nitrogen Dioxide	8			N		29-36
	Channel Status	1			N	Ø,1,2,3	29
	Channel Ø2	2			N	Always Ø2	30-31
	Polarity	1			N	+ or -	32
	Value	4			N		33-36
8	Sulfur Dioxide	8			N		37-44
	Channel Status	1			N	Ø1,2,3	37
	Channel Ø3	2			N	Always Ø3	38-39
	Polarity	1			N	+ or -	40
	Value	4			N		41-44
9	Methane	8			N		45-52
	Channel Status	1			N	Ø,1,2,3	45
	Channel Ø4	2			N	Always Ø4	46-47
	Polarity	1			N	+ or -	48
	Value	4			N		49-52
10	Total Hydrocarbon	8			N		53-60
	Chennel Status	1			N	Ø,1,2,3	53
	Channel Ø5	2			N	Always Ø5	54-55
	Polarity	1			N	+ or -	56
	Value	4			N		57-60
11	Total Oxidants	8			N		61-68
	Channel Status	1			N	Ø,1,2,3	61
	Channel Ø6	2			N	Always Ø6	62-63
	Polarity	1			N	+ or -	64
	Value	4			N		65-68

TOTAL NUMBER OF CHARACTERS

File Layout

X	INPUT		OUTPUT	SYSTEM NAME	DATE
	REFERENCE		INTERMEDIATE		

PROGRAM NUMBER(S)	FILE NAME	RECORD NAME

SEQUENCE	FREQUENCY	DISPOSITION	NO OF RECORDS PEAK: NORMAL:

DEVICE DESCRIPTION	PRINT/PUNCH DOCUMENT SIZE: WIDE X LONG NO OF COPIES

REMARKS

REF NO	DATA ELEMENT	NUMBER OF CHARS	DEC	BYTES	A P	AN B	SOURCE	RELATIVE LOCATION
12	Carbon Monoxide	8			N			69-76
	Channel Status	1			N		Ø,1,2,3	69
	Channel Ø7	2			N		Always Ø7	70-71
	Polarity	1			N		+ or -	72
	Value	4			N			73-76
13	Nitric Oxide	8			N			77-84
	Channel Status	1			N		Ø,1,2,3	77
	Channel Ø8	2			N		Always Ø8	78-79
	Polarity	1			N		+ or -	80
	Value	4			N			81-84
14	Soiling Index	8			N			85-92
	Channel Status	1			N		Ø,1,2,3	85
	Channel Ø9	2			N		Always Ø9	86-87
	Polarity	1			N		+ or -	88
	Value	4			N			89-92
15	DEW Point	8			N			93-100
	Channel Status	1			N		Ø,1,2,3	93
	Channel 1Ø	2			N		Always 1Ø	94-95
	Polarity	1			N		+ or -	96
	Value	4			N			97-100
16	Temperature	8			N			101-108
	Channel Status	1			N		Ø,1,2,3	101
	Channel 11	2			N		Always 11	102-103
	Polarity	1			N		+ or	104
	Value	4			N			105-108
17	V SIN θ	8			N			109-116
	Channel Status	1			N		Ø,1,2,3	109
	Channel 12	2			N		Always 12	110-111
	Polarity	1			N		+ or -	112
	Value	4			N			113-116
18	VCOS θ	8			N			117-124
	Channel Status	1			N		Ø,1,2,3	117
	Channel 13	2			N		Always 13	118-119
	Polarity	1			N		+ or -	120
	Value	4			N			121-124
19	Wind Speed	8			N			125-132
	Channel Status	1			N		Ø,1,2,3	125
	Channel 14	2			N		Always 14	126-127
	Polarity	1			N		+ or -	128
	Value	4			N			129-132
20	Wind Direction	8			N			133-140
	Channel Status	1			N		Ø,1,2,3	133
	Channel 15	2			N		Always 15	134-135
	Polarity	1			N		+ or -	136
	Value	4			N			137-140
	TOTAL NUMBER OF CHARACTERS							

File Layout

	INPUT		OUTPUT	SYSTEM NAME	DATE
	REFERENCE		INTERMEDIATE		
PROGRAM NUMBER(S)		FILE NAME		RECORD NAME	
SEQUENCE		FREQUENCY	DISPOSITION	NO. OF RECORDS PEAK: NORMAL:	
DEVICE DESCRIPTION		PRINT/PUNCH DOCUMENT SIZE	WIDE X	LONG	NO. OF COPIES
REMARKS					

REF NO.	DATA ELEMENT	NUMBER OF CHARS	DEC	BYTES	A AN P B	SOURCE	RELATIVE LOCATION
21	Additional Parameter No. 1	8			N		141-148
	Channel Status	1			N	Ø,1,2,3	141
	Channel 16	2			N	Always 16	142-143
	Polarity	1			N	+ or -	144
	Value	4			N		145-148
22	Additional Parameters No. 2	8			N		149-156
	Channel Status	1			N	Ø,1,2,3	149
	Channel 17	2			N	Always 17	150-151
	Polarity	1			N	+ or -	152
	Value	4			N		153-156
	Interrecord Gap						

TOTAL NUMBER OF CHARACTERS 156

(Ref. No. 6) is set to emit a constant + 1600. When the computer processes the data, if either check is outside of the prescribed limits (DVM Zero -0005 to +0005, DVM Span + 1520 to + 1680) the data for all parameters for that particular 5 minutes is declared invalid.

The system provides independent operation of each sensor. The status of each sensor is included in the 8 characters of information recorded for the 18 channels in the system. Under normal operation, the channel status Code = ø. When a sensor is undergoing calibration (either automatic or manual) the status Code = 1. The computer is programmed to check the value which is recorded for the zero calibration of each sensor. When the zero calibration is outside of preset limits, an error message is printed when the data is processed by the computer. Since the various pollutant sensors are expected to operate with essentially no drift in the zero position, the computer is not programmed to adjust for a zero drift, as in the CAMP system. Should experience with the instrument system indicate that excessive drift is occurring, it will be necessary to modify the computer program to make drift corrections.

The station operator is required to maintain a Station Activity Record (SAR) Figure 6. All pertinent information concerning station operation which is not recorded by the data acquisition system must be entered on the SAR. To begin with, the operator records the precise start and stop time of the data contained on a magnetic tape (Action = 1). In addition, the operator uses the SAR to signal the computer to invalidate

Figure 6

CINCINNATI AIR MONITORING SYSTEM

STATION ACTIVITY RECORD

Station No. I 2

Date _____ to _____

Operator _____

Sen No	Action	START						STOP						Remarks
		Mo.	Day	Yr.	Time			Mo.	Day	Yr.	Time			
3 4	5	6 7	8 9	10 11	12 13	14 15		16 17	18 19	20 21	22 23	24 25		

Sensor No. 00 = All

Action 1 = Duration of Data
2 = Invalid Data
3 = No Data
4 = Calibration

segments of data when it is apparent that some sensor (or the entire station) is malfunctioning (Action = 2).

3.3.2 Intermittent Data

Measurements of the 24-hour ambient concentration of suspended particulates, sulfur dioxide, and nitrogen dioxide, are being made at nearly 40 stations throughout the Cincinnati AQCR. Under normal circumstances, measurements are made on an every six day basis. During periods of potentially high air pollution and actual episodes, samples may be collected on a daily basis.

Intermittent data is sent to the computer center for processing on a monthly basis. New data is stored on a disk file for rapid access, for a period of three months, after which the data is transferred to the permanent data file which is maintained on magnetic tape.

In developing the procedures for handling intermittent data, two possibilities were considered; (1) recording laboratory measurements which would be converted to concentration values by a computer program, and (2) recording actual concentrations computed manually by the laboratory analyst. The decision was made to use the latter. It was agreed that spurious values are most likely to be detected when the analyst records the actual concentrations for a sample.

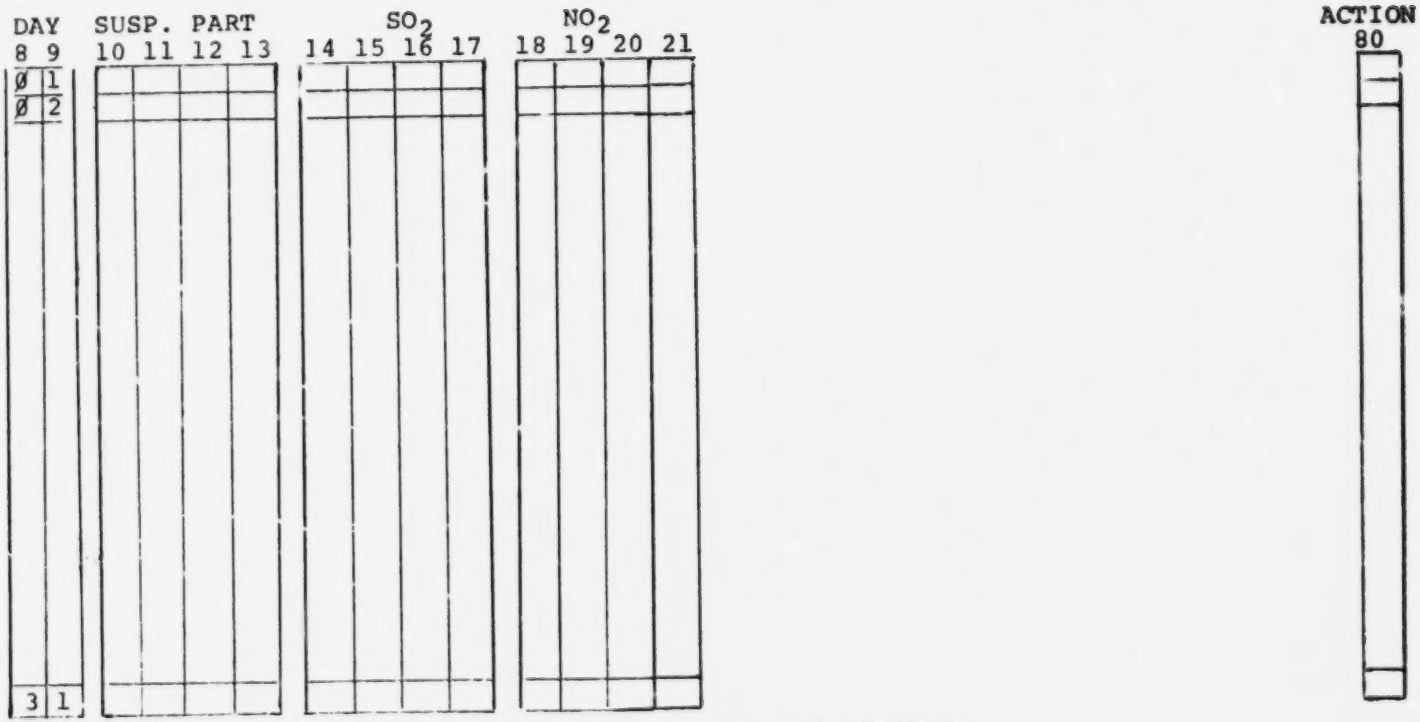
Because of the decision to process the intermittent data by the computer on a monthly basis, a monthly data record form was developed (Figure 6-A). Data for the three pollutants previously mentioned from one station is recorded on a single page. The form can easily be expanded to accommodate additional

FIGURE 6-A CINCINNATI AIR MONITORING SYSTEM
24-HOUR SAMPLES
CONCENTRATION IN $\mu g/m^3$

CA) NO. 1

STATION NO. 2 3

YEAR 4 5 MONTH 6 7



ACTION: 1 = NEW DATA 2 = CORRECT DATA NOW ON FILE

pollutants if necessary. At the end of the month, the data forms for all stations are forwarded to the computer center for keypunching.

As the intermittent data is read by the computer, the position of the data on the punched card signifies the pollutant. Pollutant codes are then attached to the data as they are stored in the data file.

4.0 OUTPUT DATA FORMATS

4.1 SURVEY USERS REQUIREMENTS

The only contact most users have with an air quality data system are the various data listings, reports, statistical summaries, etc., which the system is capable of providing. Thus, users will tend to judge the value of the system on the basis of its responsiveness to their needs. In order to satisfy the requirements of the users, it is important that most of these requirements be known at the time the system is being designed.

A necessary first step in determining user requirements is to identify the potential users of the system. While the list of users may vary somewhat from one agency to another, it will generally include the following:

A. Individuals and groups within the Control Agency.

B. Other governmental agencies at the local level.

C. State and Federal Agencies such as Department of Health and Environmental Protection Agency.

D. News media.

E. University and other researchers.

F. Trade associations.

G. Conservation groups.

H. Private citizens.

Potential users of the system should be provided with detailed information about the proposed data base, including procedures for requesting the retrieval of data and samples of output formats already developed. Users should be given sufficient time to evaluate the completeness of the system in terms of their own requirements. Finally, each user should be provided an opportunity to submit specifications for additional outputs and schedules for preparation consistent with their own needs and deadlines.

4.2 CLARITY OF CONTENT

In the design of retrieval formats for air quality data, special attention should be given to keeping the information easily understandable by the user. For example, each output format should clearly identify the name of the agency, the parameter(s) presented, where and when the data was collected, the method of determination, and the units of measurement. The use of codes are to be avoided whenever possible since most readers may find it necessary to refer to other sources of information to de-code the information.

4.3 CINCINNATI FORMATS

4.3.1 Data Listings

4.3.1.1 Continuous Data - Each week, as new data is processed and added to the disk file, a tabulation of the individual 5-minute values is prepared (Figure 7). The computer program used to prepare this listing also performs a data validation function. Each 5-minute value exceeding an estimated maximum at that sampling station is flagged. Likewise, each time

Figure 7

DIVISION OF AIR POLLUTION CONTROL

CINCINNATI OHIO

STATION STN01 CHESTERDALE AVE.

METHANE HYDROCARBONS CONCENTRATION IN P.P.M.

SEPT 29 1971

MINUTES / HOURS	0	5	10	15	20	25	30	35	40	45	50	55
0	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
1	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
2	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
3	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
4	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
5	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
6	.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
7	.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
8	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
9	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
10	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
11	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
12	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
13	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
14	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
15	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
16	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
17	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
18	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
19	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
20	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
21	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
22	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
23	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X

SEPT 30 1971

0	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
1	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
2	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
3	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
4	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
5	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
6	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
7	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
8	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
9	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
10	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
11	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
12	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
13	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
14	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
15	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
16	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
17	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
18	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
19	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
20	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
21	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
22	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
23	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X

FOOTNOTES - * PERCENT CHANGE LARGE ** EXCEEDS EXPECTED HA

successive values differ by a predetermined value, a flag is pointed. Should these flagged values (and other values) be determined invalid, an appropriate transaction to delete the data from the file is initiated by the data validation clerk.

4.3.1.2 Intermittent Data - Measurements of 24-hour pollutant concentration are added to the disk file on a monthly basis. A tabulation of the data is prepared as a part of the file updating (Figure 8). Again, the primary function of this tabulation is for data validation purposes. A second computer program is used to read user request cards calling for a listing of 24-hour data for one or more pollutants at one or more stations for a selected period of time.

4.3.2 Monthly Report

When the 5-minute pollutant concentration measurements for an entire month have been processed and validated, a monthly report is prepared (Figure 9). The basic entry in this report is the hourly average of the 5-minute data. Daily averages and the monthly average are computed from the hourly averages. Because there may be missing values in the 5-minute data, it is necessary to establish a procedure for computing an average for a time period based upon some minimum acceptable number of data values. Without such a procedure, an hourly average might be based upon one 5-minute value, or a daily average might be based upon one hourly average. The purpose in computing an average is to obtain a single value which is representative of a larger body of data.

The procedure adopted for the Cincinnati aerometric data system is to require that more than 75 percent of the possible

PEDCo COMPUTER SERVICES, INC.

Cincinnati, Ohio

PREPARED: (MONTH) XX 197X

Figure 8

DIVISION OF AIR POLLUTION CONTROL

24-HOUR MEASUREMENTS

STATION ADDRESS POLLUTANT NAME CONCENTRATION IN MICROGRAMS PER CUBIC METER

DA	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	N
01	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XX
02													
03													
04													
05													
06													
07													
08													
09													
10													
.													
.													
.													
.													
.													
.													
.													
26													
27													
28													
29													
30													
31													
n													
MIN													
MAX													
AVE													

PEDCo COMPUTER SERVICES, INC.

Cincinnati, Ohio FORMAT A

Figure 9 Format A

DIVISION OF AIR POLLUTION CONTROL

CINCINNATI, OHIO

STATION XXXX YYYYYYYYYYYYYYY PARAMETER NAME UNITS OF MEASURE MONTH 197X

HR	A.M. 0	1	2	3	4	5	6	7	8	9	10	11	P.M. 12	1	2	3	4	5	6	7	8	9	10	11	MAX. DAY 5 MIN.	AVE
DA																										
1	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XX
2																										
3																										
4																										
5																										
6																										
7																										
8																										
9																										
10																										
.																										
.																										
.																										
.																										
.																										
26																										
27																										
28																										
29																										
30																										
31																										
N	XX	XX																						NN		
AV	XXXX	XXXX																						XXXX		

MONTHLY AVERAGE XXXX

NO OF VALID DATA XXX

PEDCo COMPUTER SERVICES, INC.

Cincinnati, Ohio FORMAT B

Figure 9 Format B

DIVISION OF AIR POLLUTION CONTROL

CINCINNATI, OHIO

STATION XXXX YYYYYYYYYYYYYYYY SOILING INDEX COHS PER 1000 ft MONTH 197X

HR	A.M. 0	2	4	6	8	10	P.M. 12	2	4	6	8	10	MAX 2HR.	DAY AVE
DA														
1	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X
2														
3														
4														
5														
6														
7														
8														
9														
10														
.														
.														
.														
.														
.														
26														
27														
28														
29														
30														
31														
N	XX	XX										XX		
AV	XX.X	XX.X										XX.X		

MONTHLY AVERAGE XX.X

NO. OF VALID DATA XXX

number of data values for an averaging time be valid. Thus an hourly average is computed if there are more than 9 valid 5-minute values. Likewise a daily average is computed only if more than 18 hourly averages are valid and the monthly average (as well as the averages for individual hours for the entire month) is computed only if there are more than 23 valid daily averages.

As the hourly averages are computed, they are tabulated in the format of Figure 9 and also stored in a master file. When a valid hourly average cannot be computed, a no data code (9999) is written in the master file and the entry on the tabulation is left blank.

Note: A 5-minute value recorded as zero signifies a concentration of less than a minimally detectable amount, and is handled as a valid value in computing averages (arithmetic only).

Monthly Report Formats

FORMAT A

OPTION 1: Monthly report of parameters in primary units of measurement used for all parameters except:

a. Methane Hydrocarbons
b. Total Hydrocarbons
c. Non-Methane Hydrocarbons
d. V sin θ - no report printed
e. V cos θ - no report printed

OPTION 2: Monthly report of hydrocarbons in primary units of measurement. Replace 5 min. with 6-9 A.M. Ave. Use this format for:

a. Methane Hydrocarbons

b. Total Hydrocarbons

c. Non-Methane Hydrocarbons

OPTION 3: Monthly report of gaseous pollutants with concentration reported in micrograms per cubic meter (except carbon monoxide in milligrams per cubic meter). Do not use for the three hydrocarbons.

OPTION 4: Monthly report of hydrocarbons with concentration in micrograms per cubic meter; otherwise the same as OPTION 2.

OPTION 5: Monthly report of moving averages. Compute 3-hour moving averages of SO_2 concentration in micrograms per cubic meter. Compute 8-hour moving averages of CO concentration in milligrams per cubic meter.

FORMAT B Monthly Report of Soiling Index

Finally, a monthly average is computed if valid daily averages are available for more than 23 days during the month.

The monthly report program also computes hourly average concentrations of non-methane hydrocarbons.

Hourly average non-methane: $\bar{X}_{NM} = \bar{X}_{THC} - \bar{X}_{CH_4}$

where $\bar{X}_{THC}$ = hourly average Total Hydrocarbon

$\bar{X}_{CH_4}$ = hourly average Methane

(a) Compute $\bar{X}_{NM}$ only if $\bar{X}_{THC}$ and $\bar{X}_{CH_4}$ are <u>both</u> valid.

(b) If either $\bar{X}_{THC}$ or $\bar{X}_{CH_4}$ are invalid, enter no data codes.

(c) If $\bar{X}_{THC} - \bar{X}_{CH_4}$ negative, enter no data codes.

In the computation of the 6-9 A.M. average for Total Hydrocarbons, methane and non-methane, it is necessary to consider the fact that the sampling stations operate on Standard time throughout the year. To compensate for the time change to Daylight Savings Time, the 5-8 A.M. average is calculated for the months May-October.

The monthly report program will print pollutant concentrations in parts per million (ppm) or micrograms per cubic meter (ug/m^3). Also, the program computes three hour moving average concentrations for sulfur dioxide and eight hour moving averages for carbon monoxide. A minimum of two valid hours is required for the three hour and 6 hours for the 8 hour moving averages.

This program also computes an hourly average wind direction from the hourly average $\overline{V \sin \theta}$ and $\overline{V \cos \theta}$. The procedure for determining wind direction on a 16-point compass is as follows:

Wind Direction:

a. If hourly average wind speed <0010 (1.0 mph) record ØØ17 (CALM).

b. If hourly average wind speed >0010, determine direction on 16 point compass as follows:

(1) Compute hourly average V sin θ

($\pm \overline{V \sin \theta}$ and V cos θ ($\pm \overline{V \cos \theta}$).

(2) Comp $\overline{\sin \theta}$ and $\dfrac{\overline{V \cos \theta}}{\bar{V}}$

(3) Look up value of $\overline{\sin \theta}$ in the following table, then select proper direction depending upon the algebraic sign of $\overline{V \cos \theta}$.

NOTE: Compute $\overline{V \sin \theta}$ and $\overline{V \cos \theta}$ for those 5 minutes during the hour when V sin θ and V cos θ were both valid. Hourly averages must be based upon more than 8 valid 5-minute pairs of data.

DETERMINATION OF HOURLY WIND DIRECTION

ON 16 POINT COMPASS

$\overline{\sin \theta}$	$\overline{V \cos \theta}$	
	+	-
-0100 - -0099	13	13
-0098 - -0084	14	12
-0083 - -0056	15	11
-0055 - -0020	16	10
-0019 - +0019	1	9
+0020 - +0055	2	8
+0056 - +0083	3	7
+0084 - +0098	4	6
+0099 - +0100	5	5

CONVERSION TABLE FOR WIND DIRECTION

Code	Direction
ØØØ1	N
ØØØ2	NNE
ØØØ3	NE
ØØØ4	ENE
ØØØ5	E
ØØØ6	ESE
ØØØ7	SE
ØØØ8	SSE
ØØØ9	S
ØØ1Ø	SSW
ØØ11	SW
ØØ12	WSW
ØØ13	W
ØØ14	WNW
ØØ15	NW
ØØ16	NNW
ØØ17	CALM

4.3.3 Summary Report of Intermittent Data

An annual summary of 24-hour concentration data is prepared for each pollutant (Figure 8). In addition to tabulating individual values, the program also computes monthly averages.

4.3.4 Data Analysis

A statistical analysis of the hourly, two-hour, or 24-hour data for any period of time (i.e. monthly, quarterly, annually, etc.) can be requested (Figure 10). The user prepares a request card which specifies the sampling station(s), parameter, and begin and end dates (Figure 11). A single request card can be used to retrieve data for one station or all stations in the monitoring network.

Equations and methods used to generate the statistical analysis are presented at the end of Section 4.0 (pp 52-55).

4.3.5 Submit Data to NADB

On a quarterly basis, air quality data is submitted through the State Environmental Protection Agency to the U. S. EPA for inclusion in the National Aerometric Data Bank (NADB). The hourly averages, a 2-hour soiling index, and 24-hour pollutant concentration data are written on a magnetic tape in punched card images in SAROAD format (Figures 12 and 13).

Air Quality Data Analysis

Equations Used in Computations

No. of Samples = N, the number of measurements made during the time period specified

MIN = Minimum concentration of N measurements

MAX = Maximum concentration of N measurements

Frequency Distribution-Percentile: 10, 30, 50, 70, 90, 99

There are two methods for determining the concentrations associated with the above points on the percent cumulative frequency distribution.

PEDCo COMPUTER SERVICES, INC.

Cincinnati, Ohio

Figure 10

PREPARED: (MONTH) XX 197X

DIVISION OF AIR POLLUTION CONTROL

POLLUTANT NAME | UNITS OF MEASURE | METHOD OF ANALYSIS | SAMPLING INTERVAL XX-HR

ADDRESS	FROM YR/MO/DA	TO YR/MO/DA	NO. SAMPLES	FREQUENCY DISTRIBUTION-PERCENTILE MIN	10	30	50	70	90	99	MAX	ARITH MEAN	STD DEV	GEO MEAN	GEO STDV
XXXXXXXXXXXXXX	XX/XX/XX	XX/XX/XX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX

Figure 11.
DATA ANALYSIS PROGRAM
CONTROL CARD

No.	Name	Columns
1	Station	1-2
2	Parameter	3-4
3	Year (Begin)	5-6
4	Month (Begin)	7-8
5	Day (Begin)	9-10
6	Year (End)	11-12
7	Month (End)	13-14
8	Day (End)	15-16

Note: Station Number = ØØ causes data for this parameter to be analyzed individually for all stations.

Figure 12
PUNCHED CARD IMAGE
FOR
SUBMITTING DATA TO NADB
(TIME CODE 7)

DATA ELEMENT		COLUMNS			
NO.	NAME	FROM	THRU	TOTAL	REMARKS
1	Card No.	1	1	1	Constant 2
2	State Code (See SAROAD)	2	3	2	Constant NN
3	Area (Same as City No.)	4	7	4	Numeric
4	Site Code	8	10	3	Numeric
	Zero	8	8	1	Constant Ø
	Site No.	9	10	2	Numeric
5	Agency	11	11	1	Alpha
	F = State (1)				
	G = County (2)				
	H = City (3)				
6	Project (See Code)	12	13	2	Numeric
7	Time	14	14	1	Constant 7
8	Year	15	16	2	Numeric
9	Month	17	18	2	Numeric
10	Day	19	20	2	Numeric
11	Start Hr.	21	22	2	Constant ØØ
	PARTICULATES				
12	Parameter Code	23	27	5	Constant 111Ø1
14	Method Code	28	29	2	Numeric
15	Units	30	31	2	Numeric
16	Decimal Point	32	32	2	Numeric
17	Value	33	36	4	Numeric
	NITROGEN DIOXIDE				
18	Parameter Code	37	41	5	Constant 426Ø2
19	Method Code	42	43	2	Numeric
20	Units	44	45	2	Numeric
21	Decimal Point	46	46	2	Numeric
22	Value	47	50	4	Numeric
	SULFUR DIOXIDE				
23	Parameter	51	55	5	Constant 424Ø1
24	Method Code	56	57	2	Numeric
25	Units	58	59	2	Numeric
26	Decimal Point	60	60	1	Numeric
27	Value	61	64	4	Numeric

Figure 13
PUNCHED CARD IMAGE
FOR
SUBMITTING DATA TO NADB
(TIME CODES 1 and 2)

DATA ELEMENT		COLUMNS			
NO.	NAME	FROM	THRU	TOTAL	REMARKS
1	Card No.	1	1	1	Constant 1
2	State Code (See SAROAD)	2	3	2	Numeric (Constant NN)
3	Area (Same as City No.)	4	7	4	Numeric
4	Site Code	8	10	3	Numeric
	Zero	8	8	1	Constant Ø
	Site No.	9	10	2	Numeric
5	Agency	11	11	1	Alpha
	F = State (1)				
	G = County (2)				
	H = City (3)				
6	Project (See Code)	12	13	2	Numeric
7	Time	14	14	1	Numeric 1 or 2
8	Year	15	16	2	Numeric
9	Month	17	18	2	Numeric
10	Day	19	20	2	Numeric
11	Start Hour 00 = AM 12 = PM	21	22		Numeric
12	Parameter Code (See Code)	23	27	5	Numeric
13	Method Code	28	29	2	Numeric
14	UNITS Code	30	31	2	Ø1, Ø9
15	Decimal Point	32	32	1	Numeric
16	Reading 1	33	36	4	Numeric
17	Reading 2	37	40	4	Numeric
18	Reading 3	41	44	4	Numeric
19	Reading 4	45	48	4	Numeric
20	Reading 5	49	52	4	Numeric
21	Reading 6	53	56	4	Numeric
22	Reading 7	57	60	4	Numeric
23	Reading 8	61	64	4	Numeric
24	Reading 9	65	68	4	Numeric
25	Reading 10	69	72	4	Numeric
26	Reading 11	73	76	4	Numeric
27	Reading 12	77	80	4	Numeric

(1) Sort the N individual values (X_i) such that $X_1 < X_2 < \cdots < X_n$. The concentration X_i is determined by:

$$i = \frac{p(N + 1)}{100}$$

Example: If there are N = 75 measurements, the value X_i associated with the p = 90th percentile

$$i = \frac{90\ (76)}{100}$$

$$i = 88 \text{ (rounded to nearest whole number)}$$

(2) When N is large, it may not be desirable to sort and store the array of sorted values. In this case, it is best to compute the percentiles of the cumulative frequency distribution from a frequency distribution of the individual values as follows:

Concentration	f_i
0001	f_1
0002	f_2
0003	f_3
0004	f_4
o	o
o	o
o	o
0700	o
Greater than 700	o

where f_i = the number of times the ith concentration occurred.

Basically what is done, is to set up an array of counters where each counter corresponds to an individual

concentration. As each concentration is read from the air quality data file, a count of 1 is added into the appropriate counter.

NOTE: The maximum concentrations of the pollutants now being considered should not exceed a value of 0700. In the event that any value does exceed 0700, a final counter for greater than 700 should be provided. If this counter should exceed 1, the program should indicate with an error message that it is not possible to compute a cumulative frequency distribution.

After all values have been read and the frequency distribution is completed, compute the cumulative frequency distribution. This is done by replacing each count (f_i) with the value F_i where

$$F_i = F_{i-1} + f_i \quad i = 1,2,\cdots,N$$

thus $F_1 = f_1$

$F_2 = F_1 + f_2$

$\circ$

$\circ$

$\circ$

For each percentile (p = 10, 30, 50, 70, 90, 99) compute: $F_p = \frac{p(N + 1)}{100}$

Now do a table look up and find $F_i \geq F_p$. The concentration of the pth percentile is the concentration associated with the ith interval.

let p = 30th percentile

$$F_p = \frac{p(N+1)}{100}$$

$$F_{30} = \frac{30(34)}{100} = 10$$

$$13 \geq 10$$

concentration of 30th percentile = 0003

Arith Mean = arithmetic mean

$$= \frac{1}{N} \sum_{i=1}^{N} X_i$$

Std. Dev. = standard deviation

$$= \left[\frac{N \sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i \right)^2}{N\,(N-1)} \right]^{1/2}$$

Geo. Mean = Geometric Mean

$$= \text{antilog} \frac{1}{N} \left(\sum_{i=1}^{N} \log X_i \right)$$

Geo. Std. Dev. = Geometric Standard Deviation

$$= \text{antilog} \left[\frac{N \sum_{i=1}^{N} \log^2 X_i - \left(\sum_{i=1}^{N} \log X_i \right)^2}{N(N-1)} \right]^{1/2}$$

NOTE: In the computation of the geometric mean and geometric standard deviation, a data value of zero cannot be used. Substitute for the zero value a constant equal to one-half the minimum detectable limit for the method (e.g. if the minimum detectable value is 0.01 ppm, substitute 0.005 ppm.

5.0 DATA STORAGE

5.1 STORAGE MEDIA

5.1.1 Data Record Forms

The basic storage media is the record form on which the data was initially recorded. It is generally advisable to maintain all data record forms in an easily accessible file for a period of time. Record forms for data which are maintained in a computer oriented data base can be removed to a dead storage area relatively quickly. Data for a special short term, and special studies, as well as calibration data, may be best maintained on record form in a filing cabinet. To retain confidentiality, proprietary information may be required to be held in a filing cabinet. Retrieval of large quantities of data from data record forms is inefficient.

5.1.2 Punched Paper Tape

Punched paper tape is often used as an intermediate storage media with continuous monitoring systems equipped with data logging devices. Typically the punched paper tapes are processed by a computer and transferred to some other storage media. Problems associated with tearing and aging make paper tape undesirable for permanent storage.

5.1.3 Punched Cards

Most computer systems utilize punched cards as a primary media for data input. Punched cards are not well suited for permanent storage with large data bases. First, as card volume increases, the storage space allocation for them must

also increase. Secondly, as the cards are handled repeatedly for data retrieval, some damage is inevitable which may result in card jams in the card reader. Finally, because card reading is a slow process with most computers, continual re-reading of punched cards is both inefficient and expensive.

5.1.4 Magnetic Tape

Magnetic tape is widely used as a media for the permanent storage of data. The physical space requirements for maintaining a data base on magnetic tape is less than that of most other storage media. Data that would require several hundred thousand punched cards for storage can be written on one 2,400 foot reel of 1,600 bits per inch (BPI) magnetic tape. Next, because of the high speed tape drivers now available, data can be read into the computer very rapidly. Likewise, magnetic tapes can be read repeatedly with a very low probability of damage. The sequential nature of air quality data is ideally suited to storage on magnetic tape.

Some care in the storage and handling of magnetic tape is necessary. Ideally magnetic tapes should be stored in a fireproof vault. Lacking a suitable vault, a room with controlled temperature and humidity can be used. In this situation it is advisable to maintain a second copy of each tape in a separate room, so that a fire or other calamity will not destroy the data base. Another reason for maintaining two copies of each tape is the ever present problem of human error. Even though there are means of protecting a

tape once information has been written on it, all computer centers have experienced mistakes in tape handling. The use of more than one tape is especially important with file updating, in which data must be transferred from one tape to another as new data becomes available.

5.1.5 Direct Access Storage

The ultimate in storage media are the direct access devices such as magnetic-disk, magnetic drums, and data cells. Devices such as this are ideally suited for use with a data base which is constantly being accessed. The speeds at which data are transferred into and out of direct access storage devices provide for more efficient operation from the standpoint of the use of the Central processor of the computer system. The cost of a direct access storage device is much greater than that of a magnetic tape device. Because of the differential cost, direct access storage must be justified on the basis of savings in computer processing costs.

5.1.6 Selection of a Storage Media

The use of data record forms, paper tape, and punched cards as permanent storage media for air quality data is to be avoided in all but very specialized situations. Determining whether to use magnetic tape or direct access storage requires some additional considerations.

A magnetic tape can be purchased for about $20.00. A disk, capable of storing about twice the quantity of data that can be written on a magnetic taps costs about $500.

The differential cost can be very quickly offset through the savings in computer processing cost if the data is being frequently accessed. Since air quality data are being added continually, the data base may be accessed in a daily or weekly basis for file updating. Additionally, as the new data are validated, the file must be accessed to make changes in data already on the file. After data have been on the file for a period of time, the frequency of access very quickly decreases.

The selection of the storage media is of course dependent upon the hardward configuration of the computer system. If the system has both tape and disk drives, some combination of the two should probably be considered. The system adopted by Cincinnati utilizes both magnetic disk and magnetic tape. New data are stored on a disk file for a period of 3-months, and afterwards are transferred to magnetic tape for permanent storage. This procedure takes advantage of the high speed associated with direct access storage and the lower cost of sequential storage for historical files.

5.2 INFORMATION MANAGEMENT

The term "information management," refers to the overall process of data handling, both into and out of a computer system. To aid in the understanding of the data management processes, a discussion of some of the pertinent terminology is presented below.

5.2.1 Data Storage

In discussing data storage a differentiation is made between units of storage and units of data. Basically, units of storage relate to the computer hardware, whereas units of data are dependent upon the software programs and the user's requirements for accessing data in storage. The basic unit of addressable storage in the main storage (i.e. core) of the computer is the 8-bit byte. A byte can hold one character of information or two decimal digits. The basic unit of secondary storage is a sector on a disk, or the portion of a magnetic tape between gaps.

In main storage, a unit of data is termed a field, which consists of a fixed or floating point number, a packed decimal number, or a series of one or more bytes. Related fields, when grouped together, form a logical record (e.g. a logical record might be the necessary identification information and the 5-minute data values for one hour, for a specific pollutant). The interchange of information between main storage and secondary storage is by blocks composed of one or more logical records. A block is referred to as a physical record in secondary storage.

The effectiveness of a data management system is determined by the way in which data storage facilities are used. For example, on a magnetic tape there is a fixed length gap of 3/4" between each logical record. The use of

long physical records minimizes the portion of the total tape which is taken up by record gaps.

5.2.2 Record Formats

Logical records are grouped to form a file. The individual records can be constructed according to one of four formats; fixed length records, variable length records, undefined length records, and spanned records. The first two are most commonly used and are discussed below.

The size (i.e. number of characters or bytes) of a fixed length record is constant for all records in a file. They may be stored blocked or unblocked. Blocking of logical records is often used to create large physical records to minimize number of record gaps in the file. Computer programmers tend toward the use of fixed length records because they are more easily handled in sorting and computation.

Variable length records provide for more flexible storage of data in secondary storage. It is necessary to include the record length (l) as part of the logical record, preceding the data. Variable length records tend to maximize the efficiency of secondary storage.

In a situation where missing data may occur, space must be provided for the missing data in a fixed length record. The use of variable length records permits the use of a smaller record size and usually results in a more efficient use of secondary storage.

5.3 STRUCTURE OF THE CINCINNATI FILES

5.3.1 Use of Secondary Storage Media

The system being implemented by the City of Cincinnati makes use of both disk and magnetic tape as secondary storage media. Data from the continuous air monitors are added to the data base on a weekly basis, whereas intermittent data are entered on a monthly basis. As the new data are entered, they are processed and stored on disk for up to 100 days. During the initial storage period, the data are readily accessible for corrections and immediate reporting purposes. At monthly intervals the older data on the disk file are transferred to magnetic tape for permanent storage.

The method of utilizing both disk and tape storage provides both the efficiency of immediate access by the computer and the less expensive cost per unit of long term storage associated with magnetic tape.

5.3.2 Record Format - Disk File

The records in the disk file are all fixed length. The decision to use fixed length records was based upon software associated with the computer and the ways in which the data were to be accessed. There are three types of records used with the disk files (i.e. a file information record, data pointer records, and data records).

5.3.2.1 File Information Record - The 7200 byte file information record contains the identification of all sampling stations and parameters are being measured. Elements included in this record are as follows:

NAME	DESCRIPTION	NO. OF BYTES
APC File	Title of File	10
	CONTINUOUS DATA	
smax	Maximum number of continuous stations = 4	2
pmax	Maximum number of continuous parameters = 18	2
emax	Maximum number of pointer records per pollutant per station = 23	2
dmax	Maximum number of days in disk = 100	2
cr.date	Creation data - beginning of 100 day period	4
update	Date of last update	4
nsta	Number of continuous stations in use	2
npar	Number of parameters in use	2
last data	Disk address of last piece of data	4
hrly addr	Disk address where hourly averages start	4
	INTERMITTENT DATA	
smax	Maximum number of intermittent stations = 40	2
pmax	Maximum number of intermittent parameters = 11	2
update	Date of last update	4
nsta	Number of intermittent stations in use	2
npar	Number of intermittent parameters in use	2
int. data	Disk address where the intermittent data starts	4
	SAROAD	
STATE	SAROAD code for Ohio	2
Area	SAROAD code for Cincinnati	4
		60
	LOCATION OF CONTINUOUS STATIONS	
site	Sampling station number	3
agey	Agency code number = H	1

NAME	DESCRIPTION	NO. OF BYTES	
proj	SAROAD Project Classification, Ø1, Ø2, Ø3	2	
list	16 Switches-Parameters Measured	2	
station address	Address of first sampling station	24	
	Repeat above for second station	32	
	Repeat above for third station	32	
	Repeat above for fourth station	32	
			128
	IDENTIFICATION OF CONTINUOUS PARAMETERS		
param id	SAROAD Code for first parameter	5	
sti	SAROAD Code for time interval = Ø	1	
meth	SAROAD Code of method	2	
unit	SAROAD Code for units	2	
dec l.	Code for decimal point location	2	
st. hr.	Start Hour	2	
p.c	Parameter Control (see note)	2	
param name	First parameter name	24	
	Repeat previous 40 bytes for 2nd through 18th parameters	680	
			720
	LOCATION OF INTERMITTENT STATIONS		
site	Sampling station number	3	
agcy	Agency code number = 4	1	
proj	SAROAD project classification, Ø1, Ø2, Ø3	2	
list	11 Switches-Parameters Measured	2	
station address	Address of Sampling Station	24	
	Repeat previous 32 bytes for 2nd through 40th station	1248	
			1280
	IDENTIFICATION OF INTERMITTENT PARAMETERS		
param id	SAROAD Code for first parameter	5	
st i	SAROAD Code for time interval = 7	1	
meth	SAROAD Code for method	2	
unit	SAROAD Code for unit	2	

NAME	DESCRIPTION	NO. OF BYTES
dec 1	Code for decimal point location	2
st. hr	Start hour	2
Parameter name	First parameter name	24
	Repeat previous 38 bytes for 2nd through 11th parameter	380
		418
	Format code arrays	4594
		7200

Note: The two byte parameter control is used to identify changes in the parameter list between the 5-minute and hourly average data as follows:

a. Enter the parameter two digit parameter code (Table 3 Process Continuous Data Program) if 5-minute and hourly averages are stored for the parameter.

b. Enter ØØ for parameter (Sensor) number 11 = Vsin θ and 12 = VCos θ since no hourly averages are stored.

c. Enter 11 for 17th parameter. This will cause hourly wind direction to be stored as the 11th parameter in the hourly average file.

d. Enter 12 for the 18th parameter. This will cause hourly non-methane hydrocarbons to be stored as the 12th parameter in the hourly average file.

5.3.2.2 Data Pointer Record - Two types of data pointer records are used to maintain an inventory of the Aerometric Data Disk File. Separate data pointer records are maintained for the continuous and intermittent data files. A pointer record is written each time new data is added to the file. It is assumed that updates will be on a weekly basis. On this basis 14 updates would occur in 100 days. The file structure allows a maximum of 23 updates during a 100 day period. A 240 byte pointer record is used for each continuous parameter at each station. One 7200 character pointer record is used for all of the stations with intermittent data.

NAME	DESCRIPTION	NO. OF BYTES
	CONTINUOUS DATA	
num	Number of pointer fields in record 1st Pointer Field-Station 1 Parameter 1	2
trk	Track number	2
rec	Record number	2
yr	Year	1
day	Day	2
hr	Hour	1
lgth	Length of data record pointed	2
	Repeat previous 10 bytes for 2nd through 23rd pointer fields for Station 1 - Parameter 1	
	Blank	8
		240

Note: Hourly average data is accessed directly by calculating a displacement from the disk file starting address which is maintained in the file information record.

NAME	DESCRIPTION	NO. OF BYTES	
	INTERMITTENT DATA		
	Station 1 Parameter 1		
trk	Track number	2	
byte	Number of Bytes	2	
yr	Year	1	
day	Day of Year	2	
hr	Hour	1	
lgth	Length of Data Record Pointed	2	
	Repeat above 10 bytes for Station 1 Parameter 2 through Station 40 parameter 11	4390	
			4400
	Blank		3800
			7200

5.3.2.3 Data Records - Three different types of data records are used; (1) continuous data - 7200 bytes per record, (2) hourly data - 5400 bytes per record and (3) intermittent data - 204 bytes per record. A description of each record format is presented below.

Continuous Data

The 5-minute parameter values are written in 7200 byte records each containing the data for 257 hours plus 4 unused bytes at the end of the record. The data for a single hour requires 28 bytes used as follows:

	No. of Bytes
Year	1
Day	2
Hour	1
Value Minute ØØ	2
Value Minute Ø5	2
.	.
.	.
.	.
Value Minute 55	2

Hourly Data

The hourly averages compute for each parameter are written in 5400 byte records. The record contains the hourly data for 100 days. Each day requires 54 bytes used as follows:

	No. of Bytes
Constant Ø	1
Year	1
Day	2
Value Hour ØØ	2
Value Hour Ø1	2
.	.
.	.
.	.
Value Hour 23	2
Max 5 Min Value	2

Since soiling index is retained on the hourly average file as 2-hour values, these values are recorded in the positions for Hour ØØ, Ø2, Ø4, etc. Enter no data codes for hours Ø1, Ø3, Ø5, etc.

Intermittent Data

The 24-hour pollutant concentration data is written in 204 byte records. The record contains up to 100 data values. The 204 bytes are used as follows:

	No. of Bytes
Constant Ø	1
Year	1
Day (Begin)	2
Value Day 1	2
Value Day 2	2
.	
.	
.	
Value Day 100	2

ANNUAL AEROMETRIC DATA FILE

RECORD FORMAT

Field No.	Data Element	No. of Characters	Position In Output Record	Type of Data
1	Blank	2	1	Blank
2	Agency Code	1	3	Alpha,Always 4
3	Project Code	2	4	Num, Ø1, Ø2, Ø3
4	State Code	2	6	Num, Saroad Code
5	Area Code	4	8	Num, Saroad Code
6	Site Code	3	12	Num
7	Parameter Code	5	15	Num, Saroad Code
8	Method Code	2	20	Num, Saroad Code
9	Unit Code	2	22	Num, Saroad Code
10	Time Code	1	24	Num, Saroad Code
11	Start Hour	2	25	Num
12	Year	2	27	Num
13	Month	2	29	Num
14	Day	2	31	Num
15	Decimal Point Code	1	33	Num
16	Data Field 1	4	34	Num
17	Data Field 2	4	38	Num
18	Data Field 3	4	42	Num
.		.	.	.
.		.	.	.
.		.	.	.
46	Data Field 31	4	154	Num
47	Blank	3	158	Blank

PROJECT CODE

Long-term surveillance codes

01	Population-oriented surveillance
02	Source-oriented ambient surveillance
03	Background surveillance

Figure 14.

5.3.3 Record Format

At monthly intervals, the hourly, 2-hour, and 24-hour data are removed from the disk and transferred to a magnetic tape file for permanent storage. Data for the most current two months is retained on the disk. The 5-minute data is deleted from the disk file and is not retained on the permanent magnetic tape file.

Data is written on tape as 160 character, fixed-length, logical records. The logical records are written on tape with 20 records per block. The logical record format is identical to that used by EPA (Figure 14). The decision to use the EPA format for the historical file was influenced in part by the fact that it would simplify the submittal of data to the National Aerometric Data Bank.

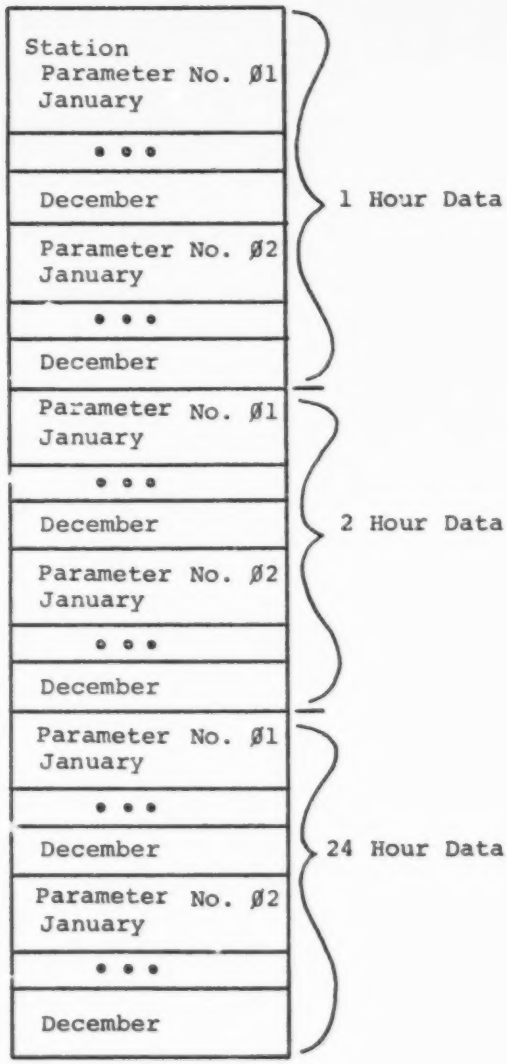
For hourly data, the start hour is normally ØØ and the 24 values for a day are stored in data fields 1 through 24. Data fields 25 through 31 are blank.

For 2-hour soiling index data, the start hour is normally ØØ and the 12 values for a day are stored in data fields 1 through 12. Data fields 13 through 31 are blank.

For 24-hour data, the start hour is normally ØØ and the values for each day are stored on the day of the month on which the sample was collected. Since 24-hour samples are normally collected on an every sixth day basis, most of the data fields in the record will be blank.

The 160 character records are written on the tape as shown in the attached Magnetic Tape Format (Figure 15).

Figure 15
Aerometric Data File
Magnetic Tape Format



A basic record in this file contains the data for one month for a parameter measured over a given sample averaging time at a given sampling station. The sort key for the file is:

Major: County
City
Site
Sample averaging time
Parameter

Minor: Month

A reel of magnetic tape will contain data from only 1 year.